An Ensemble Feature Ranking Framework for the Assessment of the Efficacy of Cervical Cancer Detection Tests and Human Papillomavirus Genotypes in the Detection of High-grade Cervical Intraepithelial Neoplasia and Cervical Carcinoma

Panagiotis Bountris, Charalampos Tsirmpas, Maria Haritou, Abraham Pouliakis, Ioannis Kouris, Petros Karakitsos, and Dimitrios Koutsouris, *Member, IEEE*

Abstract— **In most cases, cervical cancer (CxCa) develops as a result of underestimated abnormalities present in cytology via the test Papanicolaou (Pap test). Although it is the sole test that has proved its value for cancer prevention, Pap test is prone to human errors as it is performed via the microscope and requires skilled personnel. Consequently, molecular techniques have emerged in the recent years, aiming to replace Pap test or support the diagnosis as ancillary tests. These include DNA micro-arrays identifying Human Papillomavirus (HPV) subtypes, mRNA techniques such as nucleic acid based amplification or flow cytometry identifying E6/E7 oncogenes, and immunocytochemistry techniques such as overexpression of p16. These techniques are either highly sensitive or highly specific, but not both at the same time, thus no perfect method is available today. Moreover, most of the published studies compare two tests against each other, which, however, are using different population study groups, therefore it is questionable if the results can be used to perform a direct comparison between the available tests. In this paper, we adopt a sophisticated approach to assess the value of several well-known and widely applied CxCa tests in detecting high-grade cervical intraepithelial neoplasia and CxCa (CIN2+), by employing an Ensemble Feature Ranking framework. Applying the proposed framework we managed to directly compare and rank the studied CxCa tests and several HPV genotypes according to their efficacy in detecting CIN2+ lesions. The results suggest that HPV mRNA tests perform better than HPV DNA testing for the triage of abnormal Pap tests. Such findings may guide cytopathologists in a step by step basis to decide which tests to perform following an abnormal Pap test, in order to decrease uncertainty, improve accuracy and reduce time to diagnosis and resources.**

Manuscript received August 14, 2015. This study was funded by the Greek State, General Secretariat for Research and Technology - GSRT, Project "HPVGuard", Cooperation 2011-2013 (code: 11ΣΥΝ_10_250) http://HPVGuard.org.

P. Bountris is with the Biomedical Engineering Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece (phone: +30-210-7722270; fax: +30-210-7722269; e-mail: pbountris@biomed.ntua.gr).

C. Tsirmpas is with the Biomedical Engineering Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece (e-mail: htsirbas@biomed.ntua.gr).

M. Haritou is with the Institute of Communication and Computer Systems, Athens, Greece (e-mail: mhari@biomed.ntua.gr).

A. Pouliakis is with the Department of Cytopathology, "ATTIKON" University Hospital, University of Athens, Athens, Greece (e-mail: apouliak@med.uoa.gr).

I. Kouris is with the Biomedical Engineering Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece (e-mail: ikouris@biomed.ntua.gr).

P. Karakitsos is with the Department of Cytopathology, "ATTIKON" University Hospital, University of Athens, Athens, Greece (e-mail: pkaraki@med.uoa.gr).

D. Koutsouris is with the Biomedical Engineering Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece (e-mail: dkoutsou@biomed.ntua.gr).

I. Introduction

THE primary, and most valuable, test for preventing cervical cancer (CxCa) is the cytological examination via test Papanikolaou (Pap test). However, despite its application for several decades, there is lack of consensus on the optimal management of women with abnormal test results. Actually, a proportion of women having atypical squamous cells of undetermined significance (ASCUS) or low-grade squamous intraepithelial lesion (LSIL) in cytology, may have high-grade squamous intraepithelial lesion (HSIL). These women are at real risk of developing CxCa, as HSIL is the decision threshold beyond which a case is referred for treatment, whereas ASCUS or LSIL cases are only strictly monitored. The other way round, it is not infrequent that women with HSIL in cytology may have LSIL or even normal histology. These women are exposed to the physical and psychological sequelae of unnecessary colposcopies and treatments [1].

Cervical cancer is known to be caused by human papillomavirus (HPV) infection, which is the commonest sexually transmitted infection worldwide. However, the presence of HPV does not necessarily lead to disease. About 100 types of HPV virus infecting humans have been identified. Among them, at least 15 are oncogenic and can cause CxCa. Improved understanding of HPV infection and the natural history of cervical neoplasia have resulted in the addition of the HPV DNA test along with the Pap test [2].

Recently, new technologies for CxCa detection have emerged, aiming to replace Pap test or support it as ancillary tests. Some studies proposed the shift from DNA detection to mRNA identification of the viral E6/E7 oncogenes that are linked to oncogenic activation. Among them, mRNA typing with nucleic acid sequence based amplification (NASBA) [3] and Flow Cytometry [4] techniques have demonstrated promising results in increasing positive predictive value and reducing unnecessary referrals to colposcopy. At the same time, the immunocytochemical detection of genetic effects such as overexpression of p16 is

a methodology which can increase the diagnostic accuracy of Pap test [5].

Numerous published studies are attempting to clarify the role of each technique and compare it with Pap test or other methods [6-11]. The detailed analysis of these studies proves that the performance of the methods differ significantly. The techniques are either highly sensitive or highly specific, however not both at the same time, thus no perfect method is available to date. The studies are affected by the disease incidence and the prevalence of HPV infection in the population study group, therefore the individual application of one method, even if it offers a level of protection, does not reliably determine the real risk of individual women participating in screening programs.

Published studies [6-11] attempt to clarify the role of these techniques by comparing their sensitivity and specificity in the detection of high-grade cervical intraepithelial neoplasia (CIN2/3) and CxCa. Most of them compare two tests against each other, which, however, are using different population study groups, thus it is questionable if their results can be used to perform a direct comparison between the CxCa tests available today. Moreover, the fact that some of the tests have the advantage of being more specific while others have the advantage of being more sensitive, but none of them is superior over the others, makes a direct comparison between them very difficult. As a result, HPV-based tests do not have a clear role yet in screening, triage and diagnosis of cervical lesions.

In contrast to other studies, in this paper we adopt an Ensemble Feature Ranking framework in order to directly compare and rank the most widely used CxCa tests and HPV genotypes according to their efficacy in detecting CIN2+ lesions. More specifically, 49 indicators derived from the applied tests were evaluated on the basis of detecting CIN2+ utilizing 8 feature ranking algorithms. The results of these algorithms were aggregated using the Schulze aggregation method in order to produce a robust ranking of the studied tests and the HPV genotypes.

II. Materials and Methods

A. Clinical Data

Anonymized data, collected from women enrolled in a study headed by the Department of Cytopathology of the Medical School of Athens' University ("ATTIKON" University Hospital) have been used. The Institution's Ethical Review Board ("ATTIKON" University Hospital Bioethics Committee) approved the study and all experimental procedures. Participating women signed an Informed Patient Consent (ICON) form, allowing the anonymous use of their data for research purposes.

The study was carried out on ThinPrep® LBC specimens obtained from women referred for colposcopy either because (a) they had an abnormal Pap test, or (b) they volunteered to participate in the study and accepted a colposcopical examination and the application of various biomarkers on their biological material, although they had a normal Pap test (e.g. HPV positive women). In those cases that a negative Pap test was followed by negative colposcopy, no biopsy was taken and the case was considered as clinically negative.

Data were extracted from HPV lifecycle related examinations, cytological diagnoses, subsequent histological examination of biopsies (when available), visit dates, patient age, etc. These were stored into a database designed for the specific study. The patients' database includes more than 5,000 patients linked to more than 10,000 tests' series obtained from follow-up. From those, 740 cases with full tests' series were selected for this study and were further analysed (Table I). Each data series included results from the following tests:

- Pap test formulated according to the Bethesda classification (TBS2001 system) [12].
- HPV DNA test, conducted by the CLART® HUMAN PAPILLOMAVIRUS 2 test (GENOMICA), that allows simultaneous detection of 35 different high or low risk HPV genotypes [2].
- NASBA test (NucliSENS EasyQ® HPV v1.0) for identification of E6/E7 mRNA for HPV types: 16, 18, 31, 33 and 45 [3].
- PermiFlow® test (Invirion Diagnostics, LLC, Oak Brook, IL) that allows the identification of E6/E7 mRNA expression of high-risk HPV using Flow Cytometry technique (FLOW) [4], and
- Immunocytochemical expression of p16 using the CINtec® Cytology Kit [5].

Cytological findings of each patient were interpreted according to the Bethesda classification system [12] and were classified as follows (ranked): (a) within normal limits (WNL), (b) atypical squamous cells of undetermined significance (ASCUS), (c) low-grade squamous intraepithelial lesion (LSIL), (d) atypical squamous cells probably high (ASCH), (e) high-grade squamous intraepithelial lesion (HSIL), and (f) squamous cell carcinoma (SCC) or adenocarcinoma (Adeno-Ca).

A cervical biopsy was performed if Pap test revealed ASCUS or higher cytological categories (ASCUS+) or there was a visible lesion upon colposcopy. For those cases having a histological outcome, the histological diagnosis was used as the golden standard. No biopsies were obtained in clinically negative cases, which are defined as cases with negative cytology and negative colposcopy, as in these cases it is not ethical to obtain biological samples for histological examination. The three-tiered Cervical Intraepithelial Neoplasia (CIN) grading system was used for histological diagnosis. The categories, ranked according to severity, are: (a) without evidence of malignancy (Negative), (b) cervical intraepithelial neoplasia grade I (CIN1), (c) grade II or III (CIN2 or CIN3), and (d) squamous cell carcinoma (SCC) or adenocarcinoma (Adeno-Ca).

For each of the 740 cases, a feature set consisting of 49 indicators, which derived from the applied tests, was created. The result of the cytological examination has been used

TABLE I
CASES DISTRIBUTION – CYTOLOGY VS HISTOLOGY

	Cytology result					Total
	WNL	ASCUS	LSIL	HSIL	SCC/Adeno-Ca	
Histological result						
Clinically Negative	196	0	0	0	0	196 (26.5%)
Negative	35	60	22	5	0	122 (16.5%)
CIN 1	31	66	142	22	0	261 (35.3%)
CIN 2/3	3	13	27	93	0	136 (18.4%)
SCC/Adeno-Ca	0	1	2	7	15	25 (3.4%)
Total	265 (35.8%)	140 (18.9%)	193 (26.1%)	127 (17.2%)	15 (2%)	740

according to the Bethesda system. The results of the HPV DNA test examination were expressed as 35 individual variables (either positive or negative), one for each HPV DNA genotype. For the NASBA test we used the result for each individual HPV type (16, 18, 31, 33 and 45). The result of the flow cytometry test (FLOW) was used as positive or negative, as was for the result of the immunocytochemical expression of p16. Additionally, other variables expressing the HPV DNA test results were added, for instance, the existence of high-risk (HR) or low-risk (LR) types was expressed as either positive or negative.

B. Feature Selection and Ranking

In machine learning, feature selection is the process of selecting a subset of features for use in model construction [13]. Minimizing the number of features used to develop a model is important in order to reduce computation complexity, processing time, and to prevent overfitting. Feature selection techniques can be broadly categorized into two subcategories: Feature Subset Selection (FSS) and Feature Ranking (FR) [13]. FSS methods, also known as wrapper methods, evaluate a feature subset by running a learning algorithm; the usefulness of a feature subset is directly evaluated by the estimated accuracy of a trained classifier. In the FR approach, also known as filtering, feature selection is done independently of the learning algorithm of a classifier. Usually, a correlation or a class separability measuring criterion C(k), which quantifies how well each feature discriminates positive and negative examples, is adopted in order to rank all the features. The value of the criterion C(k) is computed for each of the features k=1,2,….m and the features are ranked in order of descending values. The l features corresponding to the l best values of the specific criterion are selected to form the best feature subset [13].

FR algorithms use statistical and information theory measures in order to rank the individual features according to their importance to a target concept. Therefore, these algorithms may also be used in order to evaluate the significance of individual features and perform comparisons between them. A ranked list of features might be interesting by itself, as for instance in the microarray analysis, where the ranked list of features is used to find correlations among top ranked features and some diseases.

Each FR method assesses the features using different measures and algorithms, thus they may produce different rankings. There have been proposed methodologies to evaluate those rankings [14-16]. However, it has been proved that better results may be achieved by combining different rankings [14-16]. Inspired from ensemble learning, where a combination of classifiers is used to improve predictive performance, ensemble feature ranking methodologies have been proposed for the combination of independent feature rankings using ranking aggregation techniques [15, 16]. The aggregation methods take as input a number of ranked lists produced by different FR algorithms and combine them in order to produce a better and more robust single ranked list.

In this paper, we employed 8 FR algorithms and we combined them using the Schulze aggregation method, as it is described in the following section.

C. Ensemble Feature Ranking Framework

The 49 indicators derived from the applied tests, were evaluated on the basis of their efficacy in detecting high-grade cervical intraepithelial neoplasia or cancer (CIN2+) using the following FR measures and methods:

1) Information Gain (IG): is a measure used as a splitting criterion in the decision tree construction [17]. Information gain measures the gain in information entropy achieved by partitioning the instances according to the values of an attribute. It thereby indicates the effectiveness of an attribute (feature) in classifying the training data. In this case, feature ranking is taking place by sorting (in descending order) the features according to their IG values.

2) Gain Ratio (GR): is also a measure used as decision tree splitting criterion [17]. It is a modification of IG, which was introduced by Quinlan in the C4.5 algorithm [17] in order to avoid overestimation of multi-valued features. The gain ratio normalizes the information gain score of splitting on an attribute by the entropy of this attribute. Like IG, feature ranking is taking place by sorting the features according to their GR values.

3) Symmetric Uncertainty (SU): is also a measure based on the IG measure. It normalizes the information gain score by the feature entropy times the class entropy. This value is multiplied by two in order to be rescaled into the range of zero and one [18].

4) Chi-Squared statistic (χ^2): evaluates each feature individually by measuring the chi-squared statistic with respect to the class [13].

5) T-Test statistic: evaluates each feature individually by measuring the t-test statistic with respect to the class [13].

6) Area under ROC curve (AUC): The area between the empirical Receiver Operating Characteristic (ROC) curve and the random classifier slope has been proposed as a class separability measuring criterion [13]. As presented in [13], it measures the overlap between the probability density functions describing the data distribution of a feature in two classes. This criterion serves as a measure of the class discrimination capability of a specific feature. A large AUC value corresponds to small overlap between classes, thus it corresponds to better class discrimination.

7) mRMR algorithm: minimum redundancy - maximum relevance feature selection framework; a mutual information based method which tries to select the most relevant features with respect to the target and minimize the redundancy among those selected features simultaneously [19].

8) ReliefF algorithm: It measures the quality of features according to how well their values distinguish between neighbours instances belonging to different classes [20].

Each of these methods assesses the features using different measures, thus they produce different rankings. In order to produce a more robust ranking, we combined the different rank orderings using ranking aggregation [16]. For this purpose, we adopted the Schulze aggregation method (Schwartz Sequential Dropping), which, as shown in [16], leads to significantly better results than any other ranking method, either individual or aggregated.

The Schulze method is a voting method and more specifically a Condorcet method. In other words, it is an election method that in a pairwise comparison the preferred candidate has the property of a majority rule. In the first phase of Schulze method, for each pair F*x* and F*y* of features we count how many rankings rank F*x* over F*y* and how many rankings rank F*y* over F*x*. If the first number is larger, then F*x* defeats F*y*; if the second number is larger, F*y* defeats F*x*; if both numbers are equal then there is a tie. Based on the previous calculation, a graph is constructed by adding an edge from F*x* to F*y* vertex whenever F*x* defeats F*y*. This process results to a graph with at least one cycle or single element, which is not defeated by other options. The group of all these elements forms the Schwartz set. In the second phase of the method, the appropriate edges are removed iteratively in order to break up all the existing cycles. For each iteration and in all cycles, the edge connecting F*x* to F*y* is removed if the number of rankings where F*x* is over F*y* is minimal, taking into consideration the set of edges in the specific cycle. If there are more than one edges with the same number of votes, they are all removed at once. This procedure is repeated until the Schwartz set contains no more cycles. After all cycles are broken, the Schwartz set comprises of isolated nodes only and the corresponding options are the winners. If we remove these nodes from the graph, a new group of winners can be computed. Features are ranked in the aggregated ranking according to the winner sequence they appear. It should be only one winner per round and in the case of more than one, the ties are broken randomly.

III. Results

Table II presents the 15 most significant features ranked according to their efficacy in detecting CIN2+ lesions. With regard to the oncogenic potential of several HPV genotypes, the 15 most important were ranked as presented in Table III.

TABLE II
RANKING OF MEDICAL TESTS AND BIOMARKERS ACCORDING TO THEIR EFFICACY IN DETECTING CIN2+

Rank	Medical Tests and Biomarkers
1	Pap test
2	PermiFlow® test (FLOW)
3	NASBA test (positive/negative)
4	p16
5	HR HPV
6	NASBA 16
7	HPV DNA test (positive/negative)
8	HPV 16
9	NASBA 31
10	NASBA 33
11	HPV 33
12	NASBA 18
13	HPV 31
14	HPV 51
15	HPV 52

TABLE III
RANKING OF HPV GENOTYPES ACCORDING TO THEIR POTENTIAL FOR CIN2+ DEVELOPMENT

Rank	HPV Genotypes	Characterization according to [2]
1	HPV 16	Very High Risk
2	HPV 33	Very High Risk
3	HPV 31	Very High Risk
4	HPV 51	High Risk
5	HPV 52	High Risk
6	HPV 62	Low Risk
7	HPV 6	Low Risk
8	HPV 58	High Risk
9	HPV 45	Very High Risk
10	HPV 18	Very High Risk
11	HPV 35	High Risk
12	HPV 26	High Risk
13	HPV 43	Low Risk
14	HPV 66	High Risk
15	HPV 53	High Risk

IV. Conclusions and Discussion

Test Papanicolaou is viewed as the most successful CxCa test [21], if repeatedly applied. However CxCa is not eliminated even in countries with well-organized CxCa screening programs. Nowadays there are available numerous options for the application of biomarkers in the triage of abnormal Pap smears [6-11]. Several studies report that the CxCa tests are either highly sensitive or highly specific, but not both at the same time, and that a gain in sensitivity is at the expense of specificity and vice versa [6-11]. Therefore, a direct comparison of these methods is difficult to be performed. Applying the proposed framework, it was possible to directly compare and rank the most widely used CxCa tests and biomarkers.

The results of the proposed framework (Table II) indicate that the Pap test is the most significant test, like other studies also report [6-11]. However, it should be performed by highly skilled personnel, experienced and well trained. Additionally, the specimen type (i.e. conventional or liquid based) may have an important role in diagnostic accuracy [6-11]. The second and third tests (FLOW and NASBA) are related to HPV lifecycle, i.e. when mRNA of the virus is expressed, meaning that infection has progressed. The fact that FLOW is ranked higher than NASBA (expressed as positive of negative) could be attributed to the inherent characteristics of the two methods; i.e. FLOW detects mRNA of all HR HPV subtypes while NASBA only for the most high risk (16, 18, 31, 33 and 45). Regarding the fourth test (p16), it is well known that the majority of SCCs of the cervix express p16 [5]. An interesting finding is that the HPV DNA test performed worse than the Pap test, the HPV mRNA tests and p16. This result suggests that maybe it is better to perform HPV mRNA testing or p16 for the triage of an abnormal Pap test. Similar results have also been reported from some studies [4-8, 22]. However, in contrast to these, the results of this study suggest that the ancillary tests should be performed with a specific order.

Concerning the ranking of HPV genotypes (Table III), according to the literature data the most oncogenic subtypes are HPV-16 and HPV-18. HPV DNA subtype 16 was ranked as the most important (Table III), similarly for its mRNA (see NASBA 16 in Table II), thus international scientific research data are confirmed. However DNA of HPV-18 was ranked lower than other “not so dangerous” subtypes. This may be attributed to the fact that prevalence of the virus subtypes is not similar in all geographical areas. An additional interesting finding is that the HPV genotypes 31 and 33 are ranked very high, as some studies report [10].

The outcomes of the study may have an important role for the everyday practice of a cytopathology laboratory. According to the ranking, ancillary tests can be performed in specific order, and may guide pathologists in a step by step basis on which tests to perform following an abnormal Pap test, in order to decrease uncertainty, improve accuracy, and reduce diagnosis time and use of resources, either human or financial.

References

[1] "NHS Cervical Screening Programme Statistical Bulletin (England 2012-13)," Department of Health, Ed., ed. England, 2013.

[2] J. J. Gomez-Roman, C. Echevarria, S. Salas, *et al.*, "A type-specific study of human papillomavirus prevalence in cervicovaginal samples in three different Spanish regions," *APMIS,* vol. 117, pp. 22-7, 2009.

[3] H. L. Smits, B. van Gemen, R. Schukkink, *et al.*, "Application of the NASBA nucleic acid amplification method for the detection of human papillomavirus type 16 E6-E7 transcripts," *J Virol Methods,* vol. 54, pp. 75-81, Jul 1995.

[4] R. Narimatsu and B. K. Patterson, "High-throughput cervical cancer screening using intracellular human papillomavirus E6 and E7 mRNA quantification by flow cytometry," *Am J Clin Pathol,* vol. 123, pp. 716-23, May 2005.

[5] R. Klaes, A. Benner, T. Friedrich, *et al.*, "p16INK4a immunohistochemistry improves interobserver agreement in the diagnosis of cervical intraepithelial neoplasia," *Am J Surg Pathol,* vol. 26, pp. 1389-99, Nov 2002.

[6] J. Cuzick, M. Arbyn, R. Sankaranarayanan, *et al.*, "Overview of human papillomavirus-based and other novel options for cervical cancer screening in developed and developing countries," *Vaccine,* vol. 26 Suppl 10, pp. K29-41, Aug 19 2008.

[7] C. Kottaridi, S. Tsiodras, A. Spathis, *et al.*, "Clinical performance of human papillomavirus E6, E7 mRNA flow cytometric assay compared to human papillomavirus DNA typing," *Anal Quant Cytol Histol,* vol. 33, pp. 305-10, Dec 2011.

[8] A. Trope, K. Sjoborg, A. Eskild, *et al.*, "Performance of human papillomavirus DNA and mRNA testing strategies for women with and without cervical neoplasia," *J Clin Microbiol,* vol. 47, pp. 2458-64, Aug 2009.

[9] C. Pileggi, D. Flotta, A. Bianco, *et al.*, "Is HPV DNA testing specificity comparable to that of cytological testing in primary cervical cancer screening? Results of a meta-analysis of randomized controlled trials," *Int J Cancer,* vol. 135, pp. 166-77, Jul 1 2014.

[10] E. Bernard, M. Pons-Salort, M. Favre, *et al.*, "Comparing human papillomavirus prevalences in women with normal cytology or invasive cervical cancer to rank genotypes according to their oncogenic potential: a meta-analysis of observational studies," *BMC Infect Dis,* vol. 13, p. 373, 2013.

[11] J. Cuzick, L. Cadman, D. Mesher, *et al.*, "Comparing the performance of six human papillomavirus tests in a screening population," *Br J Cancer,* vol. 108, pp. 908-13, Mar 5 2013.

[12] M. R. Henry, "The Bethesda System 2001: an update of new terminology for gynecologic cytology," *Clin Lab Med,* vol. 23, pp. 585-603, Sep 2003.

[13] S. Theodoridis and K. Koutroumbas, "Pattern recognition," 4th ed London: Elsevier: Academic Press, 2009, ch. 5.

[14] I. Slavkov, B. Zenko, and S. Dzeroski, "Evaluation Method for Feature Rankings and their Aggregations for Biomarker Discovery," *Journal of Machine Learning Research,* vol. 8, pp. 122-135, 2010.

[15] C. Sarkar, S. Cooley, and J. Srivastava, "Robust Feature Selection Technique using Rank Aggregation," *Appl Artif Intell,* vol. 28, pp. 243-257, 2014.

[16] R. C. Prati, "Combining feature ranking algorithms through rank aggregation," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, 2012, pp. 1-8.

[17] J. R. Quinlan, *C4.5 : programs for machine learning*. San Mateo, Calif.: Morgan Kaufmann Publishers, 1993.

[18] I. H. Witten and E. Frank, *Data mining : practical machine learning tools and techniques*, 2nd ed. Amsterdam ; Boston, MA: Morgan Kaufman, 2005.

[19] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J Bioinform Comput Biol,* vol. 3, pp. 185-205, Apr 2005.

[20] M. Robnik-Šikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Machine Learning,* vol. 53, pp. 23-69, 2003/10/01 2003.

[21] H. F. Nauth, *Gynecological cytology*. New York: Thieme, 2007.

[22] M. Persson, K. M. Elfstrom, S. Brismar Wendel, *et al.*, "Triage of HR-HPV positive women with minor cytological abnormalities: a comparison of mRNA testing, HPV DNA testing, and repeat cytology using a 4-year follow-up of a population-based study". *PloS one*, vol. 9(2), pp. 856-858, 2014.